

# ACCOUNTING FOR MISSING DATA IN SPARSE WAVELET REPRESENTATION OF OBSERVATION ERROR CORRELATIONS.

Vincent Chabot

Maëlle Nodet, Arthur Vidard

Météo-France  
CNRM, F31000 Toulouse, France

Inria, Université de Grenoble, CNRS  
LJK, F38000 Grenoble, France

## 1 Introduction

One of the problems in numerical weather prediction is the determination of the initial state of the system. Indeed, the true state of the atmosphere and ocean, at a given moment and in all points of space, are not accessible. In order to retrieve an optimal initial condition one uses the so called data assimilation methods that combine information from observations, model equations and their respective error statistics.

Since the late 70s, in numerical weather prediction, a dominant source of information comes from many satellites. Errors associated to such data are highly correlated in space, which can be detrimental if this is not properly accounted for. However their density in space allows for the efficient use of multi-scale transformation, which in turn permit a cheap but good approximation of said error statistics representation. The drawback of such approach is that the angling of missing data may not be trivial, the aim of this paper is to propose possible solutions to overcome this problem.

## 2 General formulation of variational data assimilation

Let  $\mathcal{M}$  be a dynamical model describing the evolution of the state variable  $X$ :

$$\left\| \begin{array}{l} \partial_t X(X_0, \mathbf{x}, t) + \mathcal{M}(X(X_0, \mathbf{x}, t)) = 0 \\ X(X_0, \mathbf{x}, t_0) = X_0 \end{array} \right. \quad (1)$$

Let  $Y(t)$  be (partial) observations of this state variable. The aim of data assimilation is to estimate an optimal initial condition  $X_0^a$  (often called analysed state) so that it is not far from the first guess  $X_0^b$  (in general coming from a previous forecast), and that the model trajectory  $X(X_0^a, \mathbf{x}, t)$  is close to the observations  $Y(t)$ . This is done by defining  $X_0^a$  as the minimum of the cost function:

$$\begin{aligned} J(X_0) &= J_b(X_0) + J_o(X_0) \\ &= \frac{1}{2} \|X_0 - X_0^b\|_{\mathcal{V}}^2 + \frac{1}{2} \int_{t_0}^{t_f} \|Y(t) - \mathcal{H}(X(X_0, \mathbf{x}, t))\|_{\mathcal{O}}^2 dt, \end{aligned} \quad (2)$$

where  $\mathcal{V}$  is the model state space,  $\mathcal{O}$  the observation space and  $\mathcal{H} : \mathcal{V} \mapsto \mathcal{O}$  the observation operator. Usually, in variational data assimilation, the minimisation is done using a gradient descent type algorithm and the gradient is computed using adjoint methods as advocated by [1].

Typically in data assimilation one uses the Mahalanobis distance  $\|\cdot\|_{\mathcal{V}}^2 = \|\cdot\|_{\mathbf{B}}^2$  and  $\|\cdot\|_{\mathcal{O}}^2 = \|\cdot\|_{\mathbf{R}}^2$  with

$$\|X\|_{\mathbf{R}}^2 = X^T \mathbf{R}^{-1} X \quad (3)$$

where  $\mathbf{R}$  and  $\mathbf{B}$  are the observation and background error covariance matrices respectively.

The choice of  $\mathbf{B}$  and  $\mathbf{R}$  is crucial in data assimilation, since it will drive the way information is spread and how redundancy of error is delt with. So far a strong research effort has been targeted on the background error statistics representation either using multi scale decomposition ([2], and references therein), recursive filters ([3]) or general diffusion models ([4]). In this paper, we mostly focus on the observation error statistics representation,  $\mathbf{B}$  is constructed using the general diffusion approach.

The  $\mathbf{R}$  matrix, however, has mostly been assumed diagonal. The main reason is that it simplifies greatly its management. Indeed the number of observation is in general non constant over time which prevents from using the same  $\mathbf{R}$  matrix at each assimilation cycle. As a consequence, at each assimilation cycle, a new  $\mathbf{R}$  matrix should be formed and inverted. Whithout speaking of the formation of the  $\mathbf{R}$  matrix, it size makes its storage and its inversion very difficult.

### 3 Satellite observation error statistics representation

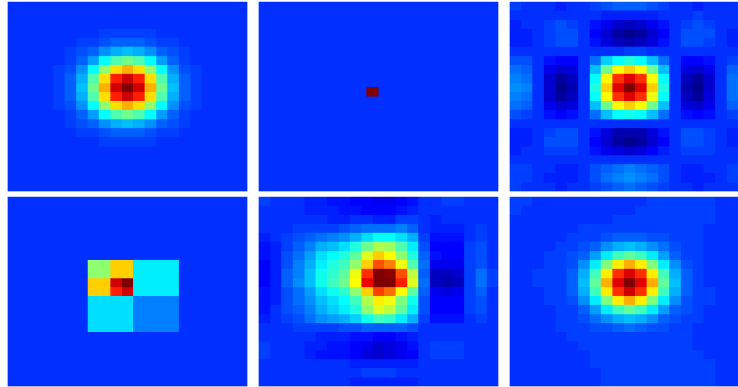
Satellite observations are in general dense in space (and time), which leads to a huge amount of data. Moreover it makes their error statistics representation difficult but in the mean time allows for the use similar approaches that the ones used for the background error representation. In particular [5] proposed to use multi-scale decomposition. The number of observation making the handling of the full  $\mathbf{R}$  matrix intractable they proposed to perform a change of variable and retain the diagonal approximation in this new space. This approach allow to account for some of the spatial correlation.

Assuming as in [6] that we can create a model equivalent (denoted  $Y = \mathcal{H}(X)$ ) to  $Y^{obs}$  the observed satellite data, one requires the definition of a distance between observed images and model output to perform data assimilation.

$$J_o(X_0) = \sum_{t_0}^{t_f} \|W(\mathcal{H}(X) - Y^{obs}(t))\|_{\mathbf{R}_w}^2 \quad (4)$$

where  $W$  is the chosen multi-scale decomposition operator. In [5] three such decompositions are studied namely a Fourier, a wavelet and a curvelet decomposition. In this paper only the wavelet decomposition is studied and the chosen wavelet decomposition is the classical 8 vanishing moments Daubechies and curvelets are described in [7].

Working in a wavelet space enable to reasonably well approximate the covariance matrix by its diagonal, making its storage and the computation of its inverse far cheaper.



**Fig. 1:** Effect of the multiplication by a Dirac impulse of the full  $\mathbf{R}$  matrix (top left) and, (from left to right, top to bottom) of diagonal approximations in pixel space ( $W = I$ ), curvelet , wavelet (Harr), wavelet (Daubechies 8) and Fourier spaces.

Figure 1 shows, for a gaussian isotropic noise, the effect of the multiplication of a Dirac impulse by  $\mathbf{R}$  (top-left) and diagonal approximations of  $\mathbf{R}$  for several  $W$ . One sees that the classical approximation is a really poor approximation while some change of variables allows for representing most of the spatial correlation. The optimal choice of this change of variable depends on the error statistics one wish to represents (see [5] for a more in-depth study). In the following we will only deal with wavelet transforms since they are a good trade-off between flexibility in representing different noise and simplicity.

## 4 Accounting for missing data

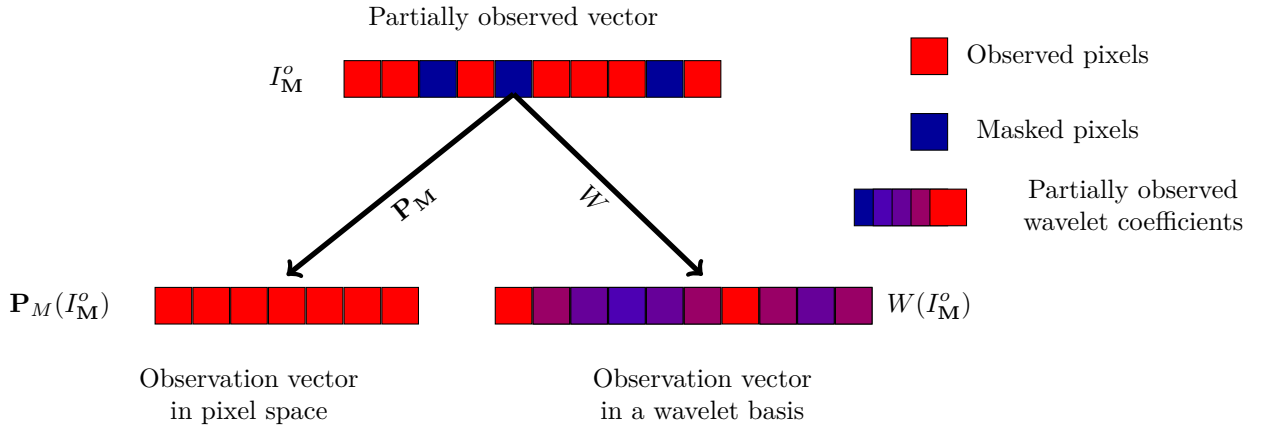
The main idea is to construct a "climatological"  $\mathbf{R}$  matrix which take into consideration the correlation model and adapted to the available datas. When using a diagonal approximation in the pixel space, such an approach is straightforward. Indeed it is just a matter of applying a projection  $\mathbf{P}_M$  to observation vector into the subset of observed pixels to get an equivalent cost function term.

$$J_{pix}^o = (\mathbf{P}_M(Y^{obs} - \mathcal{H}(X)))^T \mathbf{R}_{pixObs}^{-1} (\mathbf{P}_M(Y^{obs} - \mathcal{H}(X))) \quad (5)$$

The corresponding  $\mathbf{R}_{pixObs}$  matrix is the projection of the  $\mathbf{R}_{pix}$  matrix.

$$\mathbf{R}_{pixObs} = \mathbf{P}_M \mathbf{R}_{pix} \mathbf{P}_M^T. \quad (6)$$

When using a multi scale transformation to model (and take into account) the error correlation this is more difficult. Indeed each pixel coefficient is used in the computation of several wavelet coefficients. On the contrary each wavelet coefficient is based on several pixels. As a consequence when some pixels are missing this results in wavelet coefficient "partially" observed (see figure 2)<sup>1</sup>.



**Fig. 2:** Working in a wavelet space when the image is partially observed lead to work in larger space (but bearing the same information).

One way to account for missing data when using a wavelet transform is to tweak the diagonal coefficients based on the way their corresponding support is observed. This however is not trivial nor unique. Indeed, there are two conflicting effects of missing values. On the one hand a missing pixel does not have signal nor error, so in a sense it should reduce the error variance of the multi scale coefficients that represent it. On the other hand there are more discontinuities in the signal and the small scale coefficient are perturbed, all this should lead to inflate the error statistics.

In order to account for these two opposite effects, one can adjust the variance corresponding to the

<sup>1</sup>One can also note that  $\mathbf{R}_W$  is no longer invertible.

coefficients whose support are partially masked. For instance it can take the form:

$$\sigma_{occ}^2 = (\sigma^2 + \beta\sigma_m^2) I^2 \quad (7)$$

with  $0 \leq \beta \leq 1$  and  $0 \leq I \leq 1$ .

In order to give some possible expression of  $\beta$  and  $I$  it is necessary to recall some basics of wavelet decomposition. The principle is, at each scale, to separate the signal into a coarse approximation at that scale and finer details. Both are represented through respective coefficients ( $c$  for coarse,  $d$  for details) computed from one scale  $j$  to the next thanks to the following convolutions

$$c^{j-1}[n] = \sum_p h[p - 2n]c^j[p]$$

$$d^{j-1}[n] = \sum_p g[p - 2n]d^j[p]$$

the choice of  $g$  and  $h$  depending on the chosen type of wavelet.

Going back to equation (7)

1. On the one hand  $\beta$  corresponds to a variance inflation term, it is multiplied by  $\sigma_m^2$  the mean variance in the pixel space, corresponding to the wavelet coefficient variance when the error is decorrelated in space (and homogeneous). So  $\beta\sigma_m^2$  can be seen as a damping term toward decorrelated error signal and this especially for fine scale wavelet coefficients when the error is "smooth". When the error is "smooth", this damping term is due to the wavelet vanishing moments properties not being fully respected (because of missing data). One way to simulate this behaviour is to set  $\beta$  as

$$\beta = \left| \frac{\sum g^{occ}[p - 2n]}{\sum |g^{occ}[p - 2n]|} \right|$$

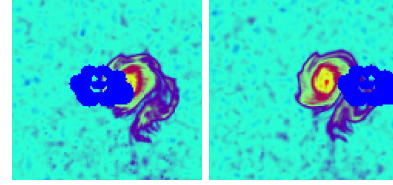
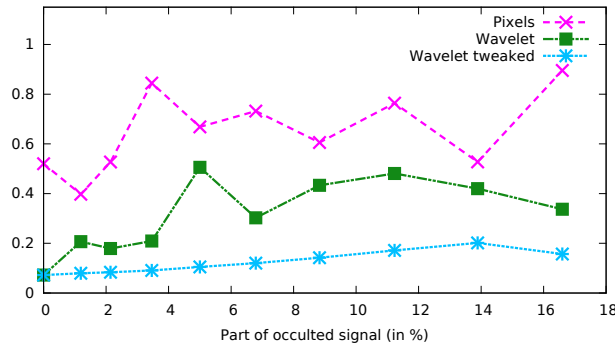
where  $g^{occ}[p - 2n] = g[p - 2n]$  if pixel  $p$  is observed and 0 otherwise.

2. On the other hand,  $I$  is a variance deflation factor. It is link to the percentage of available information in a given wavelet coefficient corresponding support. Ranging from 1 if it is fully observed to 0 if it is fully masked. In this article,  $I$  is simply the ratio of number of observed pixels over the total total number of pixels used to compute a wavelet coefficient.

## 5 Results

Twin experiments have been performed in order to validate the proposed approach. The experimental framework mimics the drift of a vortex on a turntable. The vortex is made visible thanks to the addition of a passive tracer (see [5] for more details). In the following we aim at retrieving the initial condition (the velocity field and the free surface) from an observed image sequence of passive tracer.

Figure 3 presents the residual error with respect to the occultation size. One can see that there is some benefit to account for the fact that error are correlated. Indeed, even if the variance are prescribed as in the unocculted case (green curve in Fig. 3), the results in a wavelet space are better than in the pixel space and this whatever the occultation size. The error at the end of the assimilation process is nonetheless important with respect to the unocculted case and this even if the occultations are small. This is no longer the case when the impact of missing pixels on wavelet coefficients's variance is considered (cyan curve in Fig. 3). In this case, the results are close to those obtained without occultations when they are small.



Exemples of images occulted.  
5% of the signal is missing  
Left : the first observation.  
Right : the last observation.

**Fig. 3:** Ratio between analysed state RMSE and background RMSE for data corrupted by a strong additive noise spatially correlated ( $\text{SNR} = 14.8$  dB) with respect to the percentage of pixel unobserved. The variable of interest is  $v$ .

## 6 References

- [1] Francois-Xavier Le Dimet and Olivier Talagrand, “Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects,” *Tellus A*, vol. 38A, no. 2, pp. 97–110, 1986.
- [2] Loïc Berre and Gérald Desroziers, “Filtering of Background Error Variances and Correlations by Local Spatial Averaging: A Review,” *Monthly Weather Review*, vol. 138, no. 10, pp. 3693–3720, Oct. 2010.
- [3] James R Purser, “The filtering of meteorological fields,” *Journal of climate and applied meteorology*, vol. 26, no. 12, pp. 1764–1769, 1987.
- [4] Anthony T Weaver and Philippe Courtier, “Correlation modelling on the sphere using a generalized diffusion equation,” *Q.J.R. Meteorol. Soc.*, vol. 127, no. 575, pp. 1815–1846, July 2001.
- [5] Vincent Chabot, Maëlle Nodet, Nicolas Papadakis, and Arthur Vidard, “Accounting for observation errors in image data assimilation,” *Tellus A*, vol. 67, no. 0, pp. 4117–19, Feb. 2015.
- [6] Olivier Titaud, Arthur Vidard, I Souopgui, and F.-X. Le Dimet, “Assimilation of image sequences in numerical models,” *Tellus A*, vol. 62, no. 1, pp. 30–47, Jan. 2010.
- [7] Emmanuel Candès, Laurent Demanet, David Donoho, and Lexing Ying, “Fast Discrete Curvelet Transforms,” *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861, 2006.